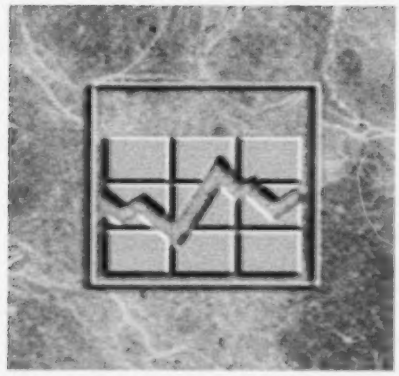
Catalogue no. 89-648-X
ISBN 978-1-100-25906-2

Longitudinal and International Study of Adults Research Paper Series

Using the Program for International Assessment of Adult Competencies direct measures of skills in the Longitudinal and International Study of Adults



by Jerry Situ

Release date: April 22, 2015

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@statcan.gc.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "The agency" > "Providing services to Canadians."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0[p] value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
[p] preliminary
[r] revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
[E] use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2015

All rights reserved. Use of this publication is governed by the Statistics Canada Open Licence Agreement.

An HTML version **is also available.**

Cette publication est aussi disponible en français.

Table of Contents

Introduction 4
Part I: PIAAC direct measures of skills in LISA 5
Literacy skills in LISA 5
Numeracy skills in LISA 6
PSTRE skills in LISA 7
Part II: Direct measures of skills in LISA versus PIAAC 8
Box 1: Treatment of literacy-related non-response in LISA 11
Conclusion 12
References 13
Annex A: Example of SAS Code used to obtain plausible values 14

Tables

Table 1
Mean literacy score for 10 plausible value sets 6
Table 2
Mean numeracy score for 10 plausible value sets 6
Table 3
Proportion of population by PSTRE levels for 10 plausible value sets 7
Table 4
Proportion of population by PSTRE levels 8
Table 5
Variance estimates by PSTRE levels 8
Table 6
Mean literacy score by age, for LISA and PIAAC, for Canada excluding territories 9
Table 7
Distribution of literacy score by age group and level, for LISA and PIAAC, for Canada excluding territories 10
Table 8
Regression results of literacy on age groups, for LISA and PIAAC, for Canada excluding territories 11

Introduction

In this paper, we demonstrate the use of the Program for International Assessment of Adult Competencies (PIAAC) direct measures of skills that are available in the Longitudinal and International Study of Adults (LISA) data.

The PIAAC direct measures of skills are available in the LISA data because the two surveys had coordinated collection. The PIAAC direct measures of skills are reflected in a set of 10 plausible values (PVs) for each of the three domains measured by PIAAC: literacy, numeracy, and problem solving in technology-rich environments (PSTRE).

The first part of this study describes and demonstrates the use of the PVs to obtain group-level estimates and estimates of the standard error. The second part of this study compares the LISA estimates to the PIAAC estimates to assess differences—as only a subset of the PIAAC respondents are included in LISA.

Part I: PIAAC direct measures of skills in LISA

The PIAAC test aimed to cover a large volume of content in a limited time. As a result, not all respondents were asked all test questions in the direct assessment. "The test design for PIAAC was based on a variant of matrix sampling (using different sets of items, multistage adaptive testing, and different assessment modes) where each respondent was administered a subset of items from the total item pool." (Organisation for Economic Co-operation and Development [OECD], 2013b, p. 408)

The responses to the subset of test items are combined with other background information (provided by the respondent) and model parameters to produce a set of 10 plausible values (PVs). These PVs can be used to produce group-level estimations of proficiency values in literacy, numeracy, and PSTRE.

Wu and Adams (2002) describe plausible values in a simple way as follows:

> The simplest way to describe plausible values is to say that plausible values are a representation of the range of abilities that a student might reasonably have. (...). Instead of directly estimating a student's ability θ,..., a range of possible values for a student's θ, with an associated probability for each of these values is estimated. Plausible values are random draws from this (estimated) distribution for a student's θ.

A single set of plausible values can be used during exploration of the data; however, to get the best estimate of the statistic and properly capture the variability in the proficiency values, all 10 sets of plausible values should be used—following the instructions provided in the OECD documentation.

OECD (2013b, Chapter 18.3) provides the following steps to calculate any estimate T of the proficiency values θ using PVs and to calculate an estimate of the variance of T:

1. Using the first vector of plausible values for each respondent, evaluate T as if the plausible values were the true values of θ. Denote the result T_1.
2. In the same manner as in step 1 above, evaluate the sampling variance of T, or $Var(T_1)$, with respect to respondents' first vectors of plausible values. Denote the result Var_1.
3. Carry out steps 1 and 2 for the second through all 10 vectors of plausible values, thus obtaining T_u and Var_u for u=2, . . .,10.
4. The best estimate of T obtainable from the plausible values is the average of the 10 values obtained from the different sets of plausible values:

$$T_. = \frac{\sum_u T_u}{10}$$

5. An estimate of the variance of T is the sum of two components: an estimate of $Var(T_.)$ obtained as in step 4 and the variance among the T_us:

$$Var(T_.) = \frac{\sum_n Var_u}{10} + (1 + \frac{1}{10}) \frac{\sum_u (T_u - T_.)^2}{10 - 1}$$

> The first component in $Var(T_.)$ reflects uncertainty due to sampling from the population; the second component reflects uncertainty because the respondents' proficiencies θ are only indirectly observed... (p. 457)

Literacy skills in LISA

Following the above steps from OECD (2013b), the estimated average of the literacy proficiency values is shown below. Notice that the statistic calculated using only 1 plausible value is a close approximation of the best estimate using all 10 plausible values; however, the variance of the estimate is underestimated with only 1 plausible value because the variability corresponding to the uncertainly of the respondents' proficiency is not reflected.

Step 1 to 3: Calculating the statistic for each of the 10 plausible value sets, and calculate the variance estimate for each statistic using the bootstrap weights. Table 1 shows the results. Annex A contains the SAS code used to obtain the values in this table.

Table 1
Mean literacy score for 10 plausible value sets

Plausible value set i	T_i	Variance of T_i
1	272.2	0.40
2	271.3	0.38
3	271.9	0.38
4	271.7	0.41
5	272.0	0.41
6	271.9	0.39
7	272.0	0.38
8	272.2	0.37
9	271.7	0.38
10	272.0	0.39

Source: Author's calculations based on Longitudinal and International Study of Adults data.

Step 4: Calculate the best estimate:

$$T_{\cdot} = \frac{\sum_u T_u}{10} = 271.9$$

Step 5: Calculate the estimate of the variance of $T_{\cdot}$:

$$\begin{aligned} Var(T_{\cdot}) &= \frac{\sum_n Var_u}{10} + \left(1 + \frac{1}{10}\right)\frac{\sum_u (T_u - T_{\cdot})^2}{10 - 1} \\ &= 0.39 + \left(1 + \frac{1}{10}\right)\frac{0.68}{10 - 1} \\ &= 0.47 \\ SD(T_{\cdot}) &= 0.69 \end{aligned}$$

Numeracy skills in LISA

Numeracy skills are calculated in a similar way as literacy skills. The estimate for the average numeracy proficiency value is shown below.

Step 1 to 3: Calculating the statistic for each of the 10 plausible value sets. Table 2 shows the results.

Table 2
Mean numeracy score for 10 plausible value sets

Plausible value set i	T_i	Variance of T_i
1	264.0	0.48
2	262.8	0.48
3	263.2	0.49
4	263.3	0.49
5	263.6	0.48
6	263.3	0.49
7	263.7	0.45
8	262.9	0.46
9	263.6	0.47
10	263.4	0.48

Source: Author's calculations based on Longitudinal and International Study of Adults data.

Step 4: Calculate the best estimate:

$$T_{\cdot} = \frac{\sum_u T_u}{10} = 263.4$$

Step 5: Calculate the estimate of the variance of $T_{\cdot}$:

$$Var(T_{\cdot}) = \frac{\sum_n Var_u}{10} + \left(1 + \frac{1}{10}\right)\frac{\sum_u (T_u - T_{\cdot})^2}{10 - 1}$$

$$= 0.48 + \left(1 + \frac{1}{10}\right)\frac{1.36}{10 - 1}$$

$$= 0.64$$

$$SD(T_{\cdot}) = 0.80$$

PSTRE skills in LISA

PSTRE proficiency values require additional consideration than that given to literacy and numeracy proficiency values because a group of respondents did not do the PSTRE assessment. The PSTRE assessment was not done if the person had insufficient computer skills, or the person opted to do a paper-and-pencil-based assessment, or the person did not do the computer assessment for literacy-related reasons. As a result, there are missing values for PSTRE proficiency values that are not addressed through imputation or weighting—as their characteristics are different from those that did complete the assessment.

Statistics Canada (2013, p. 19) recommends two approaches to conveying the volume of missing values:

1. Displaying the proportion of the population with a proficiency value when reporting any statistics calculated on the PSTRE values.
2. Displaying the proportion with missing values when reporting on the full distribution of the population by proficiency levels.

The example below shows the calculation of the proportion of the population in the different proficiency levels[1] of PSTRE.

Step 1 to 3: Calculating the statistics for each of the 10 plausible value sets. Table 3 shows the results.

Table 3
Proportion of population by PSTRE levels for 10 plausible value sets

Plausible value set i	$T_{i,\text{below level 1}}$	Variance of $T_{i,\text{below level 1}}$	$T_{i,\text{level 1}}$	Variance of $T_{i,\text{level 1}}$	$T_{i,\text{level 2}}$	Variance of $T_{i,\text{level 2}}$	$T_{i,\text{level 3}}$	Variance of $T_{i,\text{level 3}}$
1	0.157	0.000026	0.308	0.000043	0.288	0.000037	0.064	0.000012
2	0.155	0.000024	0.309	0.000041	0.283	0.000034	0.071	0.000013
3	0.162	0.000025	0.296	0.000036	0.289	0.000036	0.070	0.000012
4	0.159	0.000026	0.303	0.000041	0.289	0.000037	0.066	0.000012
5	0.156	0.000024	0.309	0.000043	0.279	0.000035	0.073	0.000013
6	0.155	0.000026	0.307	0.000044	0.287	0.000043	0.068	0.000012
7	0.157	0.000027	0.300	0.000039	0.295	0.000039	0.065	0.000012
8	0.158	0.000025	0.303	0.000040	0.284	0.000036	0.072	0.000012
9	0.159	0.000026	0.300	0.000040	0.295	0.000037	0.064	0.000010
10	0.163	0.000025	0.299	0.000042	0.285	0.000038	0.070	0.000012

PSTRE: Problem solving in technology-rich environments
Source: Author's calculations based on Longitudinal and International Study of Adults data.

1. OECD (2013, Chapter 18) classifies proficiencies values into different levels to provide more descriptive information about group proficiencies. High levels are associated with higher proficiencies.

Step 4: Calculate the best estimate. Table 4 shows the results.

Table 4
Proportion of population by PSTRE levels

PSTRE level	Proportion of population
$T_{\text{below level 1}}$	0.158
$T_{\text{level 1}}$	0.303
$T_{\text{level 2}}$	0.288
$T_{\text{level 3}}$	0.068

PSTRE: Problem solving in technology-rich environments
Source: Author's calculations based on Longitudinal and International Study of Adults data.

Step 5: Calculate the estimate of the variance of T_r. Table 5 shows the results.

Table 5
Variance estimates by PSTRE levels

PSTRE level	Variance of T_r	Standard deviation of T_r
$T_{\text{below level 1}}$	0.000034	0.0058
$T_{\text{level 1}}$	0.000065	0.0081
$T_{\text{level 2}}$	0.000064	0.0080
$T_{\text{level 3}}$	0.000024	0.0049

PSTRE: Problem solving in technology-rich environments
Source: Author's calculations based on Longitudinal and International Study of Adults data.

Missing PSTRE proficiency values are directly observed—unlike the proficiency values which are indirectly observed. Thus, estimates related to the population without PSTRE proficiency can be estimated using only one set of plausible values instead of all ten sets—as there is no uncertainty related to the fact that the PSTRE proficiency value is not available. The proportion without PSTRE proficiency values is 0.183, and its standard error is 0.0054.

Part II: Direct measures of skills in LISA versus PIAAC

LISA included a subsample of PIAAC respondents. The PIAAC sample covered all provinces and territories in Canada; it included supplementary samples for youths, official language minorities, immigrants, Métis, and Aboriginals in select provinces. The final number of respondents obtained in PIAAC is 27,285.

The subsample of PIAAC selected for LISA was selected using systematic sampling of the PIAAC sample excluding the supplementary samples and sampled households in the territories. PIAAC respondents also had to complete the questionnaire up to the point where an assessment score could be assigned to be considered a PIAAC respondent in LISA (see Box 1 for more details). The final number of PIAAC respondents in LISA with PIAAC assessment scores is 8,598. Countries involved in PIAAC were asked to have a minimum of 5,000 cases; thus, the LISA subsample is as large as the complete PIAAC sample of many other countries.

The advantage of using LISA over PIAAC is the availability of additional variables and administrative data with which proficiency scores can be analyzed. However, LISA's smaller sample size compared to PIAAC results in a higher variability in the estimate; and thus, also less power in statistical tests for significance. Literacy score estimates are provided below to give a flavour of the estimates from LISA compared to PIAAC. Numeracy and PSTRE estimates can be expected to behave in a similar fashion.

Table 6 shows the estimates of mean literacy proficiency scores by age group using LISA compared to PIAAC data. The standard error is about 20% larger for the LISA estimates compared to the PIAAC estimates.[2] Nevertheless, the LISA literacy mean estimates are reliable and also match the relative rankings observed with the PIAAC estimates.

Table 6
Mean literacy score by age, for LISA and PIAAC, for Canada excluding territories

	Literacy					
	LISA			PIAAC		
Age group	Average	Standard error	Coefficient of variation	Average	Standard error	Coefficient of variation
			%			%
16 to 24 years	273.1	1.9	0.7	275.9	1.3	0.5
25 to 34 years	284.6	1.7	0.6	285.2	1.3	0.4
35 to 44 years	278.5	1.7	0.6	279.7	1.4	0.5
45 to 54 years	265.2	1.7	0.6	268.0	1.3	0.5
55 to 65 years	259.5	1.5	0.6	260.4	1.1	0.4
Total, 16 to 65 years	271.9	0.7	0.3	273.6	0.6	0.2

LISA: Longitudinal and International Study of Adults
PIAAC: Programme for the International Assessment of Adult Competencies
Source: Author's calculations based on Longitudinal and International Study of Adults data and Programme for the International Assessment of Adult Competencies data.

Table 7 shows the distribution of literacy scores by level of proficiency by age group. As to be expected with a smaller sample size in LISA compared to PIAAC, the coefficients of variation of the estimates in LISA are higher and in some instances, unreliable. Nevertheless, the estimated proportion of people in most levels is reliable and the relative rankings between the levels matches what is observed in PIAAC.

2. Estimates of standard error for PIAAC are based on the delete-one jackknife method; while estimates of standard error for LISA are based on the bootstrap method.

Table 7
Distribution of literacy score by age group and level, for LISA and PIAAC, for Canada excluding territories

		LISA			PIAAC		
Age group	Literacy level	Proportion	Standard error	Coefficient of variation	Proportion	Standard error	Coefficient of variation
		%					
16 to 24 years	Below level 1	3.1[E]	0.8	26.1	2.6[E]	0.5	19.0
	Level 1	12.3	1.4	11.3	10.7	0.9	8.4
	Level 2	33.4	2.1	6.4	32.9	1.7	5.1
	Level 3	40.1	2.0	4.9	41.9	1.5	3.6
	Level 4	10.7	1.4	13.4	11.4	1.0	8.7
	Level 5	F	F	52.3	F	F	40.6
25 to 34 years	Below level 1	2.3[E]	0.7	30.9	1.9[E]	0.4	20.5
	Level 1	9.6	1.2	12.3	9.2	0.9	9.6
	Level 2	27.7	1.8	6.4	28.7	1.7	5.8
	Level 3	40.8	2.0	4.9	40.1	2.1	5.1
	Level 4	18.2	1.5	8.2	18.3	1.4	7.5
	Level 5	F	F	38.0	1.7[E]	0.5	27.1
35 to 44 years	Below level 1	3.0[E]	0.6	20.4	3.3	0.5	15.8
	Level 1	11.7	1.3	11.4	11.0	1.0	9.1
	Level 2	29.0	1.8	6.1	28.1	1.5	5.2
	Level 3	40.6	1.8	4.3	40.9	1.3	3.1
	Level 4	14.5	1.3	8.8	15.3	1.1	6.9
	Level 5	F	F	37.5	1.4[E]	0.4	28.9
45 to 54 years	Below level 1	5.8	0.8	14.4	5.2	0.6	12.1
	Level 1	15.5	1.4	9.1	14.6	0.9	6.1
	Level 2	33.7	1.9	5.5	33.0	1.3	4.0
	Level 3	33.1	1.8	5.5	34.5	1.4	3.9
	Level 4	11.1	0.9	8.3	11.8	0.9	7.5
	Level 5	F	F	42.6	0.8[E]	0.2	26.2
55 to 65 years	Below level 1	5.3	0.6	12.2	5.7	0.6	9.8
	Level 1	17.8	1.1	6.2	17.1	0.9	5.4
	Level 2	38.0	1.6	4.2	37.0	1.3	3.4
	Level 3	31.4	1.5	4.7	32.1	1.1	3.4
	Level 4	7.2	0.9	12.4	7.8	0.8	10.6
	Level 5	F	F	58.7	F	F	48.9
Total, 16 to 65 years	Below level 1	4.0	0.3	8.1	3.8	0.2	6.3
	Level 1	13.5	0.7	5.0	12.6	0.5	3.6
	Level 2	32.4	0.8	2.6	32.0	0.7	2.1
	Level 3	37.0	0.9	2.4	37.7	0.7	1.8
	Level 4	12.3	0.5	4.1	12.9	0.5	3.8
	Level 5	0.8[E]	0.2	18.5	1.0	0.1	14.8

F too unreliable to be published
[E] use with caution
LISA: Longitudinal and International Study of Adults
PIAAC: Programme for the International Assessment of Adult Competencies
Source: Author's calculations based on Longitudinal and International Study of Adults data and Programme for the International Assessment of Adult Competencies data.

Table 8 shows results of a simple regression model of literacy scores with age groups. Literacy score is the dependent variable, and the independent variables are indicator variables for the age groups (with the age group 16-24 being the reference group). The regression formula is:

$$y_{literacy\ score} = a + b_1 \times Age_{25-34} + b_2 \times Age_{35-44} + b_3 \times Age_{45-54} + b_4 \times Age_{55\ or\ older}$$

The intercept reflects the mean score for the age group 16-24 and is the same value as observed in Table 6. The coefficient estimates are similar in both models. The standard errors of the estimates are higher with LISA than with PIAAC; nevertheless, both LISA- and PIAAC-based models are able to show statistical differences between the age groups.

Table 8
Regression results of literacy on age groups, for LISA and PIAAC, for Canada excluding territories

	LISA					PIAAC				
	Estimate	Variance	Standard error	t	p-value	Estimate	Variance	Standard error	t	p-value
a: Intercept	273.1	3.7	1.9	142.2	<.0001	275.9	1.6	1.3	216.5	<.0001
b1: Age(25 to 34 years)	11.5	6.7	2.6	4.4	<.0001	9.4	3.3	1.8	5.1	<.0001
b2: Age(35 to 44 years)	5.4	7.0	2.7	2.0	0.0439	3.8	3.8	1.9	2.0	0.0514
b3: Age(45 to 54 years)	-8.0	7.2	2.7	-3.0	0.0030	-7.9	2.8	1.7	-4.7	<.0001
b4: Age(55 to 65 years)	-13.6	6.1	2.5	-5.5	<.0001	-15.5	2.7	1.7	-9.4	<.0001

LISA: Longitudinal and International Study of Adults
PIAAC: Programme for the International Assessment of Adult Competencies
Source: Author's calculations based on Longitudinal and International Study of Adults data and Programme for the International Assessment of Adult Competencies data.

Box 1: Treatment of literacy-related non-response in LISA

Literacy-related non-response in the PIAAC subsample selected for LISA is treated differently than in the full PIAAC sample. OECD (2013a, p. 69) explains literacy-related non-response as a result of adults being "unable to complete the background questionnaire as they were unable to speak or read the language of the assessment, had difficulty reading or writing, or had a learning or mental disability".

Literacy-related non-response records do not have assessment scores and have minimal background information. The group is also likely to have low levels of proficiency. As a result, PIAAC does not adjust for literacy-related non-response through weight adjustments as it does for other types of non-response. Instead, these records are kept on the dataset with non-zero weights and with missing values for the variables. This allows data users to measure the size of the literacy-related non-response and make adjustments as they see fit.

LISA, on the other hand, treats literacy-related non-response in the same manner as other forms of non-response—that is, through weight adjustments. As a result, literacy-related non-respondents in the PIAAC subsample selected for LISA has zero-value PIAAC weights—as weights of other PIAAC respondents are adjusted to account for the non-response.

The different treatment of literacy-related non-response in LISA was based on the following considerations:

1. Only 0.9% of the PIAAC weighted population in the 10 provinces were literacy-related non-respondents; thus, limiting the differences resulting from different treatment of literacy-related non-response.
2. Most literacy-related non-respondents were also non-respondents to the LISA questions. Assuming data users would always be using LISA variables in conjunction with assessment scores (otherwise, the full PIAAC dataset would be a better data source), missing values in the LISA variables limit a data user's ability to make literacy-related non-response adjustments.
3. Most data users would exclude the literacy-related non-response records from their analysis if they were in the dataset. This effectively treats the non-response as completely missing at random. Addressing the literacy-related non-response through weight adjustments results in comparatively less biased estimates as the relation to the weighting and calibration variables are taken into account.

The impact of different treatments of literacy-related non-response on literacy and numeracy mean scores is demonstrated in OECD (2013a, p. 69). It indicates that complete exclusion of literacy-related non-respondents produces an upper bound of estimated proficiency scores for literacy and numeracy, while an assumed score of 85 points presents a reasonable lower bound. Using these two approaches, OECD (2013a) estimates, for Canada, the two treatments of literacy-related non-responses has a difference of 1.7 for the mean literacy score and 1.6 for the mean numeracy score.

Conclusion

This paper demonstrates the use of direct measures of literacy, numeracy, and PSTRE in LISA and provides a flavour of how it compares to estimates produced from the PIAAC dataset. We conclude that:

1. Data exploration can be based on a single set of plausible values (with variance estimations using bootstrap weights). This will produce reliable point estimates, but will underestimate the standard error as the variability resulting from indirectly observing the proficiency scores is not included.
2. The best and final estimates should be based on all ten sets of plausible values, following the directions in OECD (2013b).
3. Analysis of proficiency scores using LISA should also make use of variables available in LISA but not in PIAAC; otherwise, the PIAAC dataset should be used instead, as the larger sample size allows for more precise estimates.
4. LISA estimates of proficiency scores cannot be directly compared to published estimates from PIAAC due to different treatment of literacy-related non-response and different geographical coverage. However, reliable estimates can be produced from LISA; and, the relative rankings of proficiency estimates for subgroups and the conclusions that are drawn from them should be similar to what would be concluded from the full PIAAC dataset.

References

Organisation for Economic Co-operation and Development. (2013a). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Retrieved from: http://dx.doi.org/10.1787/9789264204256-en

Organisation for Economic Co-operation and Development. (2013b). *Technical Report of the Survey of Adult Skills (PIAAC)*. Retrieved from: www.oecd.org/site/piaac/_Technical%20Report_17OCT13.pdf

Statistics Canada. (2014). *LISA, 2012: Quick Start User Guide*. Ottawa, ON: Statistics Canada.

Statistics Canada. (2013). *User Guide: The Programme for International Assessment of Adult Competencies, 2012*. Ottawa, ON: Statistics Canada.

Wu, M.L., and Adams, R. J. (2002). *Plausible Values: Why they are important*. Paper presented at the International Objective Measurement workshop, New Orleans, LA.

Annex A: Example of SAS Code used to obtain plausible values

```
/*************************************************************************************************
The following code uses BOOTVAR v3.2 to generate estimates for the mean literacy score and the
standard error of the estimate for each of the 10 plausible value sets.
*************************************************************************************************/
libname mylib "{directory}"; /*Sets personal directory*/

/*Reading in LISA Wave 1 data*/
%LET datafid = '{directory}\LISA_MASTER_2011.txt';
data mylib.lisa_w1;
   %include '{directory}\LISA_MASTER_2011_i.SAS';
   %include '{directory}\LISA_MASTER_2011_lbe.SAS';
   if IRPW_FLG = 1; /*Keeping only ISA respondents*/
   keep PersonID IRPW PVLIT1-PVLIT10 dummy; /*Keeping only the necessary variables*/
   dummy=1; /*creating dummy variable for denominator in the ratio calculation of the average*/
run;

/*Getting PIAAC respondent person bootstrap weights (IRPW)*/
%LET datafid = "{directory}\LISA_MASTER_2011_BOOTSTRAP_irpw_w1.txt";
data mylib.lisa_bsw;
   %include "{directory}\LISA_MASTER_2011_BOOTSTRAP_irpw_w1_i.sas";
run;
/*Add final weight to bootstrap weight file because BOOTVAR requires it*/
data mylib.lisa_bsw2;
   merge mylib.lisa_bsw (IN=a) mylib.lisa_w1(keep=personid irpw);
   by personid;
   if a;
run;

/*************************************************************************************************
******************** START OF TEMPLATE PROVIDED BY BOOTVAR SOFTWARE ******************************
*************************************************************************************************/

*************************************************************************
** SPECIFY THE NAME OF THE FOLLOWING 2 DIRECTORIES (directories only): **
*************************************************************************;
      libname in1 "{directory}";
      libname out "{directory}";

*************************************************************************************
** SPECIFY THE NAME OF THE ANALYSIS FILE (CREATED IN STEP 1) (without extension):  **
*************************************************************************************;
      %let Mfile = in1.lisa_w1;

*************************************************************************
** SPECIFY THE NAME OF THE FILE CONTAINING THE BOOTSTRAP WEIGHTS:       **
*************************************************************************;
      libname in2 "{directory}";
      %let bsamp=in2.lisa_bsw2;

*************************************************************************************************
** SPECIFY, IF DESIRED, THE BREAKDOWN VARIABLE(S) (EG: PROVINCE, SEX, ETC...):                  **
**     Write the name of the breakdown variable(s) below.                                       **
**                                                                                              **
**     -  If the analysis includes all of the data in the file created in step 1, put           **
**        a dot. (%let classes =. )                                                             **
**     -  If more than one variable, leave a blank between each variable                        **
**        (%let classes=var1 var2)                                                              **
**     -  DO NOT ERASE OR COMMENT OUT THIS COMMAND                                              **
*************************************************************************************************;
      %let classes = .;


*************************************************************************************************
** SPECIFY THE FOLLOWING INFORMATION (SPECIFIC TO THE SURVEY YOU ARE USING):                    **
**    You must specify:                                                                         **
**          1- The unique identifier variable(s) (separated by a space)                        **
**          2- The final weight (variable included in the bootstrap weight file)               **
**          3- The prefix of the bootstraps weight variables                                   **
**          4- Parameter for the mean bootstrap (set to 1 for regular bootstrap)               **
```

```
**          5- The number of bootstrap weights to use (note: For testing, B must be >= 2. **
**             IT IS NECESSARY TO USE ALL THE BOOTSTRAP WEIGHTS WHEN PERFORMING THE FINAL **
**             ANALYSIS.)                                                                **
**                                                                                       **
**          - Refer to AppendixC  to obtain this information                             **
*******************************************************************************************;
      %let ident = personid;
      %let fwgt  = IRPW;
      %let bsw   = BSIRPW;
      %let R     = 1;
      %let B     = 1000;


*****************************************************************************
** SPECIFY THE DIRECTORY AND THE NAME OF THE FILE THAT CONTAINS THE MACROS  **
** (THE PROGRAM MACROE_V32.SAS IF NO MODIFICATIONS HAVE BEEN MADE BY THE USER) **
*****************************************************************************;
      %include "{directory}\SASBootvar_v32_20110704\MACROE_V32.SAS";

/*****************************************************************************/
/***                                                                       ***/
/***                              SECTION 2                                ***/
/***                                                                       ***/
/*** This section lets the user specify the different analyses of interest.***/
/***                                                                       ***/
/*****************************************************************************/

* TO OBTAIN VARIANCE ESTIMATES OF A RATIO, RUN:
  ------------------------------------------;
%ratio(PVLIT1, dummy);
%ratio(PVLIT2, dummy);
%ratio(PVLIT3, dummy);
%ratio(PVLIT4, dummy);
%ratio(PVLIT5, dummy);
%ratio(PVLIT6, dummy);
%ratio(PVLIT7, dummy);
%ratio(PVLIT8, dummy);
%ratio(PVLIT9, dummy);
%ratio(PVLIT10, dummy);

%output;  /*Displays the results on the screen.  Do not modify. */

/*****************************************************************************************************
******************** END OF TEMPLATE PROVIDED BY BOOTVAR SOFTWARE ************************************
*****************************************************************************************************/
```